



HEART DISEASE PREDICTOR SYSTEM

Sharanabasavaraj H AngadiBasavarajPatil

Associate Professor

Asst Professor

Department of Computer Science and Engg

Department of Information Science and Engg

RTE Society's Rural Engineering College,SDM Institute of Technology Ujire-India

Hulkoti- Karnataka

Email:bbpatilcs@gmail.com

Abstract— Heart disease is a major cause of morbidity and mortality in modern society. Medical diagnosis is a complicated task that should be performed accurately and efficiently and its automation would be very useful. All doctors are not equally skilled in every sub specialty and they are in many places a scarce resource. The WHO (World Health Organization) has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases. Half the deaths in the United States and other developed countries occur due to cardiovascular diseases. On the whole, heart disease is regarded as the primary reason behind deaths in adults. The term Heart ailment includes the different ailments that influence the heart. Coronary illness was the significant reason for losses in the various nations including India. It kills one individual at regular intervals in the United States. The proposed method delineates a system for automated medical diagnosis would enhance medical care and reduce costs. It is unreasonable for a typical man to as often as possible experience exorbitant tests like the ECG and subsequently there should be a framework set up which is convenient and simultaneously dependable, in anticipating the odds of coronary illness. Thus, developing an application which can predict the vulnerability of a heart disease given basic symptoms and person's age, sex, pulse rate etc.

Keywords—Heart Disease; Medical diagnosis; coronary illness; Cardiovascular disease; Heart ailment

INTRODUCTION

Heart Disease is one of the main sources of death around the world covering over a portion of the demise on the planet. As indicated by the insights, the commonness of coronary illness and stroke has expanded by over half from 1990 to 2019 in India, with an expansion saw in each state. The commitment of these infections to add up to passing and illness trouble in the nation has nearly multiplied in the previous 25 years. Coronary illness currently is the main individual reason for malady trouble in India, and stroke is the fifth driving reason. The CDC (Centres for Disease Control and Prevention) expresses that in the United States, somebody has a cardiovascular failure like clockwork. Every moment, more than one individual in the United States passes on from a coronary illness-related occasion and it costs the United States about \$200 billion every year. This absolute incorporate the expense of social insurance administrations, drugs, and lost efficiency.



Volume 8, Issue 6 – June 2020 - Pages 54-63

All heart-related sicknesses can be pre-recognized and rewarded in this way the danger of having bitten the dust from a coronary illness can be decreased. There are numerous sorts of coronary illnesses, including: (hypertension), coronary illness (respiratory failure), and cerebrovascular sickness (stroke). Hypertension, high LDL cholesterol, and smoking are key coronary illness chance variables for coronary illness and some different components are diabetes, less than stellar eating routine physical idleness, and overweight. Every one of these infection types can be pre-recognized by normally checking glucose, pulse, and so forth. This prompts the issue of significant expense of clinical charges and regular check-ups.

The framework which is created can discover the heart sicknesses chance factor by investigating clinical qualities that the client can assemble from minimal effort to keep away from the normal meet up with the specialists and medicinal services registration. The framework utilizes AI calculation to be a specific choice tree to distinguish the danger of having the coronary illness so the client can be progressively wary about their wellbeing and keep up a sound way of life. The framework gets 13 data sources, for example, ... which will be investigated with the calculation and yield a level of danger of having a coronary illness.

The framework will be accessible in two diverse temperaments of getting to which are through the web and android application. This makes the client get to information advantageously and without any problem. Besides, examining the given information and some extra information the framework will deliver the arrangement of social insurance tips that the client can catch up to keep up a reasonable and solid way of life. The client will be cautioned with the wellbeing condition which they must be generally worried about in his life and if necessary the framework will create a cautioning message to the client to make a meeting with the specialist.

LITERATURE REVIEW

Authors [1], states that the greater part of the authors in information digging groupings strategies proposed for the expectation of coronary illness, however the forecast framework didn't consider the vulnerability in the information measure. In this way, to expel the vagueness and vulnerability, authors made an analysis with fluffy methodology by acquainting a participation work with the classifier. To evacuate the vulnerability in the information, through fluffy methodology presenting an exponential enrolment work with Standard Deviation and Mean determined for the properties estimated for every individual and the resultant participation esteem is increased with the credits to expel vulnerability. The presentation investigation shows that the fluffy K-NN classifier is increasingly exact as contrasted and K-NN classifier where the K-NN classifier has the exactness factor between 40 to 50 percent and fluffy K-NN classifier scores a precision between 80-90 percent.

Vishva et.al [2] states, with the assistance of information mining strategies, specialists will have the option to make future forecasts of coronary illness. The procedures which have been utilized by the creators for enormous information are Data Mining, Machine Learning and Principal Component Analysis which is one of the most broadly utilized numerical instruments for high dimensional information examination is Principal Component Analysis (PCA). Creators [2] utilized the system of utilizing a few AI calculations which are specifically Decision Tree, Random Forest, Logistic Regression, Support Vector Machines, MLP Classifier and Navie Bayes. Results shows Logistic Regression has the most precision of 86% without PCA and Random Forest has the most noteworthy exactness of 84% with PCA.



Volume 8, Issue 6 – June 2020 - Pages 54-63

Authors [3] expect AI classification is the most operational appraisal creation strategies for this present reality and scientific circumstances Prediction precision may fluctuate restrictively on various learning methods. AI classification is the most operational evaluation creation techniques for this present reality and scientific circumstances. Ashok Kumar et.al [3] applied six AI strategies including artificial neural system (ANN), bolster vector machine (SVM), calculated relapse, k-closest neighbour (kNN), classification tree and Navie Bayes. In addition, the presentation was looked at utilizing collector employable trademark (ROC) and adjustment chart. The most elevated classification exactness of 85 % was accounted for utilizing calculated relapse with affectability and specificity of 89% and 81%, individually.

Deepthi et.al [4] essential objective was to develop a Model of Heart Disease Prediction System using information mining showing method, explicitly, Clustering using apportioning procedure. Heart disease desire system can discover and focus covered taking in related with coronary ailment from a recorded heart sickness database. It can answer complex requests for Analysis of heart disease and thusly help human administrations specialists to choose clinical choices. The creators utilized Partitioning Method, K-Means Algorithm and Feature Selection Method(PSO) for testing the most appropriate calculation for coronary illness expectation.

Authors [5] introduced a review of different models dependent on AI calculations and methods to break down their presentation. Models depend on managed learning calculations, for example, Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF) and gathering models are discovered extremely well known among the analysts. Irregular woods got most noteworthy precision of 97.7%.

Santhana et.al [6] utilized administered information mining calculations on the dataset to foresee the potential outcomes of having coronary illness of a patient, they were broke down with order models to be specific Naïve Bayes Classifier and Decision tree grouping. These two calculations are applied to the equivalent dataset so as to break down the best calculation as far as exactness. The Decision tree model has anticipated the coronary illness persistent with an exactness level of 91% and Naïve Bayes classifier has anticipated coronary illness understanding with a precision level of 87%.

Authors [7] utilized the neural system calculation multi-layer perceptron (MLP) to prepare and test the dataset. Multi-layer perceptron calculation is a regulated neural system calculation wherein there will be one layer for the information, second for the yield and at least one for concealed layers between these two layers. They utilized 7 ascribes to foresee which the clients can get to without any problem.

Authors [8] utilized seven classifiers, specifically DT, NB, MLP, RFB, SCRL, K-NN and SVM, and mixes of these classifiers, utilizing gathering learning techniques, for example, stowing, boosting and stacking, are talked about. In every situation, the exhibition is determined utilizing the standard measurements, specifically exactness, accuracy, review and F-measure. What's more, the Receiver Operation Characteristic (ROC) bend territory has been utilized to think about the exhibition of every classifier.



Volume 8, Issue 6 – June 2020 - Pages 54-63

Srabanti et.al [9] proposed a framework utilizing choice tree procedure, i.e., C4.5 calculation is joined with ANN and named as half breed DT to create the ideal outcome. The dataset is stacked, and pruning is never really clamor. The proposed strategy is applied on the clean dataset. The exactness, affectability, and specificity of individual calculation and the cross breed DT are resolved.

Authors [10] utilized four AI calculations in particular choice tree, Naïve Bayes, artificial neural system and boosting/packing/outfit. The creators made the assessment models as review, exactness, accuracy and F measure.

Ritika et.al [11] composed the paper pointing towards finishing up the most efficient method among Neural Networks, Decision Tree and Naive Bayes utilized for the expectation of coronary illness based on precision, forecast rate and time multifaceted nature. It additionally represents the system or usage devices utilized for every one of them.

Authors [12] not just gives a structure to the arrangement that velocities up the conveyance and lessens mistakes by decreasing the measure of composing code yet in addition gives the strategies including assessment and sending, stacking and change. The graphical UI of this apparatus decides the outcome with an efficient planning. Here serum cholesterol is more unmistakable than others. At the point when input datasets of various patients are applied on fast excavator apparatus, at that point the exhibition think of an exactness of about 71.43 %.designations.

PROPOSED WORK

Coronary illness is one of the significant reasons for death on the planet paving the way to half of the all-out passing every year. It is likewise a significant reason for handicap. There are various types of coronary illness. The most well-known reason for coronary illness is narrowing or blockage of the coronary conduits, the veins that flexibly blood to the heart itself. This is called coronary vein illness and happens gradually after some time. It's a significant explanation individuals have respiratory failures. Different sorts of heart issues may happen to the valves in the heart, or the heart may not siphon well and cause a cardiovascular breakdown.

To decrease the number of passing the most plausible arrangement is to foresee heart ailments before it occurs. The cutting edge world has many propelled methods of distinguishing heart ailments through specialists however it is still at an early age to anticipate coronary illness by the client itself without the consistent meetings with the specialist. The fundamental issue with the regular check-ups is it will cost a great deal of cash for the patients for arrangements and furthermore for the distinctive registration which will be endorsed by the specialists. The difficult explanation is to structure a framework that will be minimal effort and an effective plan of a framework to anticipate coronary illness. The framework ought to be easy to understand and simple to access by the clients. The framework must be equipped for anticipating coronary illness by examining the information of the patients in which the patients should contribute to the framework. The information which will be a contribution by the customers must be values that will be accessible to them with the least expense.

DESCRIPTION OF DATABASE

Define The openly accessible Cleveland coronary illness database comprises of 303 cases where the turmoil is one of four sorts of coronary illness or its absence. There are fourteen factors as shown in Table I.

Table 1. Clinical features and their description

Observation	Description
age	Age in years
Sex	Sex of person
CP	Chest Pain
trestbps	Resting blood pressure
chol	Serum cholesterol
fbs	Fasting blood sugar
restecg	Resting electrocardiographic result
thalach	Maximum heart rate achieved
exang	Exercise induced angina
oldpeak	ST depression induced by exercise relative to rest
slope	Slope of peak exercise ST segment
ca	Number major vessels colored by fluoroscopy
thal	Defect type
num (disorder)	Heart Disease

In Table 1, there are 14 attributes utilized in this technique, including 8 symbolic and 6 numeric: age (age in years), sex (male or female), Chest pain type (typical angina, atypical angina, non-angina pain, asymptomatic), trestbps (resting blood pressure), cholesterol (serum cholesterol in mg/dl), fasting blood sugar < 120 mg/dl (true or false), resting electrocardiographic results (normal or having ST-T wave abnormality), maximum pulse, exercise-induced angina (true or false), oldpeak (ST depression induced by exercise relative to rest), slope (up, flat, down), number of vessels colored by fluoroscopy (0-3), thal (normal, fixed defect, reversible defect), and num (healthy, with heart-illness)

ARTIFICIAL NEURAL NETWORK

Artificial Neural Network(ANN) is a computational system that is inspired by structure, processing model and learning ability of the biological brain. It consists of large number of very simple processing neuron-like processing elements. The processing potential of the network is stored within the inter-unit connection strengths (called weights) gained by a process of adaptation to, or learning from, a group of training patterns.

The network contains three layers: the input layer, the hidden layer and the output layer, having 13 neurons, 6 neurons and 2 neurons respectively. Figure 1 displays the framework of the model.

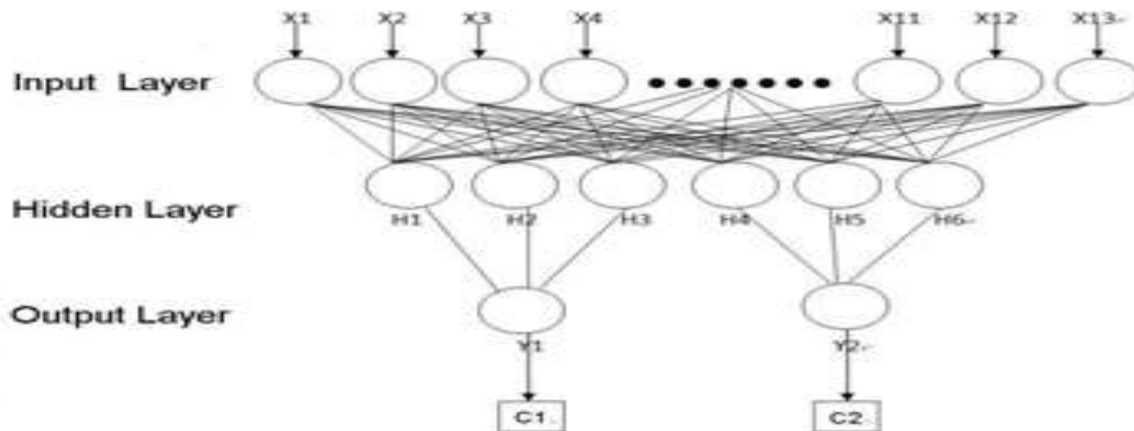


Figure 1. Framework of the ANN Model:

The model contains three layers: the input layer, the hidden layer and the output layer, having 13 neurons, 6 neurons and 2 neurons respectively.

SYSTEM DESIGN

A person with some clinical data can prognosticate the chances of getting heart disease. The person has to provide all the data obtained from the clinic and put it into the system. After processing all such data, the system displays whether he/she is would get heart disease or not. Along with this result system also provides some tips to the person for his healthy life. After getting results a person can consult a cardiologist for further treatment if required.

To have a conversation with the system, which is to provide his/her clinical data and getting results and tips, a person should have an interface. This interface work as a medium between the person and the system.

FLOWCHART

The flowchart of the proposed system design is represented in figure 2.

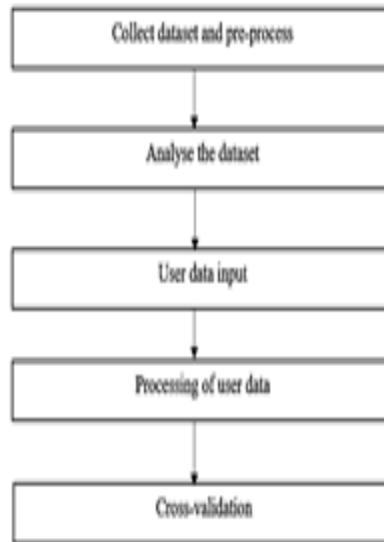


Figure 2. Flowchart of the proposed system

Interface

The system is designed using Python programming language. Python package called tkinter is used to design user interface of the system. The main functions of the system interface include:

- (1) Input clinical data section: users input 13 pieces of clinical data.
- (2) Predict button: users click the button to get the result.
- (3) Prediction result: this shows the prediction result of the provided clinical data.

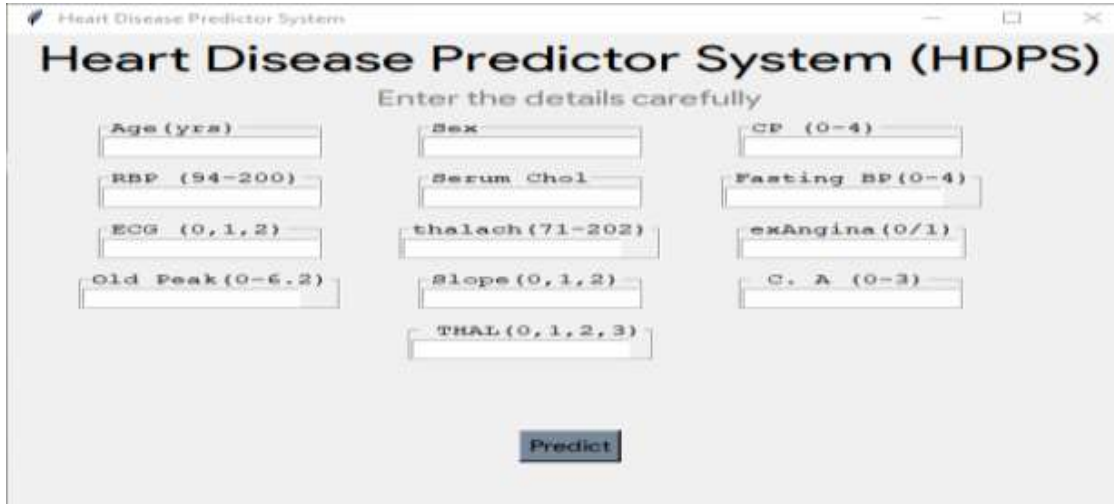
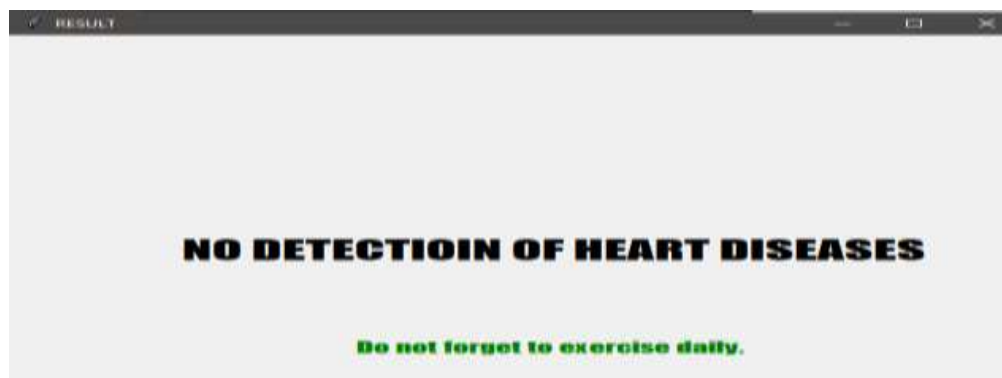


Figure 3. Heart Disease Prediction System User Interface

RESULTS AND DISCUSSION

The output is based on the clinical data provided by the user. If the user data found to be vulnerable to heart disease the system shows a message saying that the person is suffering from heart disease. If not, particular message is shown to the user.





CONCLUSION

Heart illness expectation framework can find and concentrate shrouded learning related with coronary illness from a recorded heart illness database. It can answer complex inquiries for Analysis of heart illness and along these lines help human services experts to settle on clinical decisions.

In future this work is to be reached out by utilizing other data mining algorithms and enhancement algorithms for foreseeing different illnesses from the informational collection by utilizing less number of attributes. The proposed work can be additionally improved and extended for the mechanization of Heart disease prediction. Genuine information from Health care associations and offices should be gathered and all the accessible procedures will be looked at for the ideal accuracy.

REFERENCES

- 1) V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "**Heart Disease Prediction System Using Data Mining Technique by Fuzzy K-NN Approach**", Emerging ICT for Bridging the Future, 2015, 978-3-319-13728-5, Doi: 10.1007/978-3-319-13728-5_42.
- 2) Ching-seh (Mike) Wu, Mustafa Badshah, VishwaBhagwat, "**Heart Disease Prediction Using Data Mining Techniques**", 2nd International Conference on Data Science and Information Technology, 2019, 978-1-4503-7141-4, 2019, Doi: 10.1145/3352411.3352413
- 3) Ashok Kumar Dwivedi, "**Performance Evaluation of Different Machine Learning Techniques for Prediction of Heart Disease**", The Natural Computing Applications Forum, 29, 2018, 0941-0643, May 2018, Doi: 10.1007/s00521-016-2604-1.
- 4) DeepthiVaddella, CheepirisetiSruthi, Bharath Kumar Chowdary, Somula Rama-subbareddy, Govinda. K, "**Prediction of Heart Disease Using Machine Learning Techniques**", January 2019, ISSN:0097-8043, Vol-118-Issue-1-January-2019.
- 5) V.V. Ramalingam, AyantanDandapath, M Karthik Raja, "**Heart disease prediction using machine learning techniques: a survey**", International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687, 2018.
- 6) Mr. Santhana Krishnan, "**Prediction of Heart Disease Using Machine Learning Algorithms**", 2018, 978-1-7281-1604-4, Doi: 10.1109/ICIICT1.2019.8741465.
- 7) AditiGavhane, GouthamiKokkula, IshaPandya, Prof. Kailas Devadkar, "**Prediction of Heart Disease Using Machine Learning**", 2018, 978-1-5386-0965-1, Doi: 10.1109/ICECA.2018.8474922.



Volume 8, Issue 6 – June 2020 - Pages 54-63

- 8) SourabhKulkarni, Chaitra. D Bhat, DeepaPatil, JovitaDara, "**Heart Disease Classification: A Case Study using Machine Learning and Data Mining**", National Conference on Engineering Innovations and Solutions (NCEIS - 2018), 2018, ISSN: 2456-3307
- 9) SrabantiMaji and SrishtiArora, "**Decision Tree Algorithms for Prediction of Heart Disease**", Information and Communication Technology for Competitive Strategies, 2019, https://doi.org/10.1007/978-981-13-0586-3_45.
- 10) EktaMaini, BonduVenkateswarlu, and Arbind Gupta, "**Applying Machine Learning Algorithms to Develop a Universal Cardiovascular Disease Prediction System**", Springer Nature Switzerland, 2019, https://doi.org/10.1007/978-3-030-03146-6_69.
- 11) RitikaChadha, ShubhankarMayank, "**Prediction of heart disease using data mining techniques**", Springer India, 2016, 10.1007/s40012-016-0121-0.
- 12) Pinky Bajaj, KavitaChoudhary and RenuChauhan, "**Prediction of Occurrence of Heart Disease and Its Dependability on RCT Using Data Mining Techniques**", Springer India, 2015, 10.1007/978-81-322-2247-7_86.